

# **MEMORY DEVICE WITH HIGH DIELECTRIC CONSTANT GATE DIELECTRICS AND METAL FLOATING GATES**

## **TECHNICAL FIELD OF THE INVENTION**

**[0001]** The present invention relates generally to memory devices and in particular the present invention relates to flash memory device architecture.

## **BACKGROUND OF THE INVENTION**

**[0002]** Memory devices are typically provided as internal, semiconductor, integrated circuits in computers or other electronic devices. There are many different types of memory including random-access memory (RAM), read only memory (ROM), dynamic random access memory (DRAM), synchronous dynamic random access memory (SDRAM), and flash memory.

**[0003]** Flash memory devices have developed into a popular source of non-volatile memory for a wide range of electronic applications. Flash memory devices typically use a one-transistor memory cell that allows for high memory densities, high reliability, and low power consumption. Common uses for flash memory include personal computers, personal digital assistants (PDAs), digital cameras, and cellular telephones. Program code and system data such as a basic input/output system (BIOS) are typically stored in flash memory devices for use in personal computer systems.

**[0004]** Stray capacitance in flash memory cells can cause problems. For example, the capacitance between different floating gates that are close together can cause coupling and cross-talk between the floating gates of neighboring cells. This may also have the effect of reducing memory cell performance.

**[0005]** Figure 1 illustrates a cross-sectional view of a typical prior art memory cell array. A typical cell is comprised of a silicon substrate 100. A gate insulator layer 101 is formed on top of the substrate 100. Oxide isolation areas 103 and 104 are formed between the cells. The floating gates 105 and 106 are formed between the oxide isolation areas 103 and 104. An interpoly insulator 107 is formed over the floating gates 105 and 106 prior to

forming the control gate 110 on top. The memory array is comprised of multiple rows 120 and 121 of memory cell transistors.

[0006] The capacitances that couple the various components of the array are illustrated as  $C_{A-D}$ .  $C_A$  is the row – to – row floating gate stray capacitance.  $C_B$  is the end – to – end floating gate stray capacitance.  $C_C$  is the floating gate – to – control gate coupling capacitance and  $C_D$  is the floating gate – to – substrate coupling capacitance.

[0007] The ratio of these capacitive components is determined by the geometrical dimensions of the facing surfaces constituting the capacitance and the dielectric constants of the insulator materials. The ends and sides of the floating gates are the plate areas of the stray capacitances. The dielectrics between the side and end areas are the oxide and have the same dielectric constant as the gate oxide. In the case of NAND flash memory devices, the polysilicon floating gate material is thick resulting in large surfaces on the ends and sides of the floating gates. The thick floating gate material results in greater stray capacitances.

[0008] For the reasons stated above, and for other reasons stated below which will become apparent to those skilled in the art upon reading and understanding the present specification, there is a need in the art for a flash memory cell transistor that has reduced stray capacitance and, therefore, higher performance.

## SUMMARY

[0009] The above-mentioned problems with stray capacitance and other problems are addressed by the present invention and will be understood by reading and studying the following specification.

[0010] The present invention encompasses a flash memory transistor with a high permittivity gate dielectric and metal floating gate to reduce the effect of stray capacitances. The memory transistor comprises a substrate that has a plurality of source/drain regions. The source/drain regions have a different conductivity type than the remainder of the substrate.

[0011] A metal floating gate is formed overlying the substrate. In one embodiment, the floating gate is formed over a high dielectric constant tunnel insulator. A metal oxide inter-gate insulator is formed over the metal floating gate. The inter-gate insulator and the tunnel insulator each have a dielectric constant that is greater than a dielectric constant of silicon dioxide. A control gate is formed on top of the inter-gate insulator layer. In one embodiment, the control gate is metal.

[0012] Further embodiments of the invention include methods and apparatus of varying scope.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

[0013] Figure 1 shows a cross-sectional view of a typical prior art flash memory cell transistor array showing stray capacitances.

[0014] Figure 2 shows a cross-sectional view of one embodiment of a flash memory cell transistor array of the present invention.

[0015] Figure 3 shows a block diagram of an electronic system of the present invention.

### **DETAILED DESCRIPTION**

[0016] In the following detailed description of the invention, reference is made to the accompanying drawings that form a part hereof and in which is shown, by way of illustration, specific embodiments in which the invention may be practiced. In the drawings, like numerals describe substantially similar components throughout the several views. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention. Other embodiments may be utilized and structural, logical, and electrical changes may be made without departing from the scope of the present invention. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope of the present invention is defined only by the appended claims and equivalents thereof.

[0017] In the subsequent discussion of the embodiments of the present invention, high dielectric constant (high-k) materials are considered to have dielectric constants that are greater than that of silicon dioxide (i.e.,  $k = 3.9$ ). Alternate embodiments define high-k materials with different dielectric constants.

[0018] Figure 2 illustrates a cross-sectional view of one embodiment of a NAND flash memory cell transistor array of the present invention. Each transistor is comprised of a high dielectric constant (high-k) tunnel insulator, a metal floating gate, and a high-k inter-gate insulator. Examples of gate structure embodiments of the present invention include a deposited aluminum oxide – aluminum floating gate – aluminum oxide grown by low temperature oxidation – aluminum control gate, deposited aluminum oxide – aluminum floating gate – deposited aluminum oxide – aluminum control gate, and PbO grown by oxidation of Pb – Pb – PbO grown by oxidation of Pb. These embodiments are for purposes of illustration only since the present invention is not limited to any one structure.

[0019] Referring to Figure 2, each transistor of the array is comprised of two source/drain regions 201 that are doped into a silicon substrate 200. The second source/drain region is not shown but is located along the z-axis behind the first source/drain region 201. Since the illustrated embodiment is a NAND architecture array, the second transistor row 221 behind the first row 220 shares the second source/drain region. In one embodiment, the source/drain regions are n+ regions and the substrate is p-type silicon. However, the present invention is not limited to any conductivity type.

[0020] Additionally, the present invention is not limited to NAND-type memory arrays. Alternate embodiments may use NOR architecture arrays or other types of flash memory architectures.

[0021] A high-k tunnel oxide layer 209 is formed on the substrate 200 over the source/drain regions 201 and between low dielectric constant oxide isolation areas 211 and 212. The isolation areas 211 and 212 are formed in the substrate and separate the columns of the memory array.

[0022] The metal floating gate/trapping layer 203 is formed on top of the tunnel oxide layer 209 and a high-k, metal oxide inter-gate insulator layer 207 is formed on top of the

trapping layer 203. The high-k oxide –metal floating gate – high-k oxide layers 209, 203, and 207 form a composite gate insulator under the control gate 205. The control gate 205 can be comprised of a polysilicon and/or metal. Alternate embodiments use other semiconductor materials for the control gate.

**[0023]** The rows 220 and 221 of the memory array are separated by a low dielectric constant oxide isolation material 230. Similarly, the columns of the array are separated by the low dielectric constant oxide isolation areas 211 and 212.

**[0024]** In one embodiment, the high-k tunnel dielectric 209 may be formed by deposition techniques. The inter-metal insulator layer 207 may be formed by low temperature oxidation of metals. The metal floating gate 203 can be deposited on the tunnel dielectric 209 using atomic layer deposition (ALD), evaporation, sputtering, or other techniques. These techniques for forming the various layers of the transistor are for purposes of illustration only. Alternate embodiments can use other techniques.

**[0025]** Single layers of Ta<sub>2</sub>O<sub>5</sub>, TiO<sub>2</sub>, ZrO<sub>2</sub>, Nb<sub>2</sub>O<sub>5</sub>, or similar transition metal oxides can be formed by low temperature oxidation of numerous transition metal films. Examples of such operations are discussed subsequently. These metal oxide layers can also be formed by ALD, chemical vapor deposition (CVD), and sputtering.

**[0026]** For purposes of illustration, typical composite gate insulator structures can include Ta<sub>2</sub>O<sub>5</sub> – Ta – Ta<sub>2</sub>O<sub>5</sub>, TiO<sub>2</sub> – Ti – TiO<sub>2</sub>, ZrO<sub>2</sub> – Zr – ZrO<sub>2</sub>, and Nb<sub>2</sub>O<sub>5</sub> – Zr – Nb<sub>2</sub>O<sub>5</sub>. In each of these, the metal oxide can be deposited or grown. Alternate embodiments may use other compositions.

**[0027]** The oxide growth rate and limiting thickness increases with oxidation temperature and oxygen pressure. The oxidation kinetics of a metal may, in some cases, depend on the crystallographic orientations of the very small grains of metal that comprise the metal films of the present invention. If such effects are significant, the metal deposition process can be modified in order to increase its preferred orientation and subsequent oxide thickness and tunneling uniformity. To this end, use can be made of the fact that metal films strongly prefer to grow during their depositions having their lowest free energy planes parallel to the film surface. This preference varies with the crystal structure of the

metal. Metal orientation effects, if present, would be larger when only a limited fraction of the metal will be oxidized and unimportant when most or all of the metal is oxidized.

[0028] As is well known in the art, ALD is based on the sequential deposition of individual monolayers or fractions of a monolayer in a well-controlled manner. Gaseous precursors are introduced one at a time to the substrate surface. Between the pulses the reactor is purged with an inert gas or evacuated.

[0029] In the first reaction step, the precursor is saturatively chemisorbed at the substrate surface and during subsequent purging the precursor is removed from the reactor. In the second step, another precursor is introduced on the substrate and the desired films growth reaction takes place. After that reaction, byproducts and the precursor excess are purged from the reactor. When the precursor chemistry is favorable, one ALD cycle can be performed in less than one second in a properly designed flow-type reactor.

[0030] The most commonly used oxygen source materials for ALD are water, hydrogen peroxide, and ozone. Alcohols, oxygen and nitrous oxide have also been used. Of these, oxygen reacts very poorly at temperatures below 600°C but the other oxygen sources are highly reactive with most of the metal compounds listed above.

[0031] Source materials for the above-listed metals include: zirconium tetrachloride ( $\text{ZrCl}_4$ ) for the Zr film, titanium tetraisopropoxide ( $\text{Ti}(\text{OCH}(\text{CH}_3)_2)_4$ ) for the Ti film, trimethyl aluminum ( $\text{Al}(\text{CH}_3)_3$ ) for the Al film. Alternate embodiments use other source materials.

[0032] Thin oxide films are deposited at a temperature that is high enough such that, when it is adsorbed to the substrate surface, the vaporized source material reacts with a molecular layer of a second source material or that the vaporized source material becomes adsorbed and reacts with the second source material directed to the substrate surface in the subsequent step. On the other hand, the temperature should be low enough such that thermal breakdown of the source material does not occur or that its significance in terms of the total growth rate of the film is very small. Regarding the above-listed metals, the ALD

process is carried out at a temperature range of approximately 200-600°C. Alternate embodiments use other temperature ranges.

[0033] In another embodiment of the memory transistor of the present invention, the metal floating gate and high-k metal oxide insulator layers can be fabricated using evaporation techniques. Various evaporation techniques are subsequently described for some of the high dielectric constant materials listed above.

[0034] Very thin films of TiO<sub>2</sub> can be fabricated with electron-gun evaporation from a high purity TiO<sub>2</sub> slug (e.g., 99.9999%) in a vacuum evaporator in the presence of anion beam. In one embodiment, an electron gun is centrally located toward the bottom of the chamber. A heat reflector and a heater surround the substrate holder. Under the substrate holder is an ozonizer ring with many small holes directed to the wafer for uniform distribution of ozone that is needed to compensate for the loss of oxygen in the evaporated TiO<sub>2</sub> film. An ion gun with a fairly large diameter (3 – 4 in. in diameter) is located above the electron gun and argon gas is used to generate Ar ions to bombard the substrate surface uniformly during the film deposition to compact the growing TiO<sub>2</sub> film.

[0035] A two step process in fabricating a high-purity ZrO<sub>2</sub> film avoids the damage to the silicon surface by Ar ion bombardment. A thin Zr film is deposited by simple thermal evaporation. In one embodiment, this is accomplished by electron beam evaporation using an ultra-high purity Zr metal slug (e.g., 99.9999%) at a low substrate temperature (e.g., 150° – 200°C). Since there is no plasma and ion bombardment of the substrate, the original atomically smooth surface of the silicon substrate is maintained. The second step is the oxidation to form the desired ZrO<sub>2</sub>.

[0036] The nitridation of the ZrO<sub>2</sub> samples comes after the low-temperature oxygen radical generated in high-density Krypton plasma. The next step is the nitridation of the samples at temperatures > 700°C in a rapid thermal annealing setup. Typical heating time of several minutes may be necessary, depending on the sample geometry.

[0037] Using a Pb/PbO structure, the Lead Oxide (PbO) barrier may be controllably grown on deposited lead films using either thermal oxidation or RF sputter etching in an

oxygen plasma. One processing sequence using such a thermal oxidation process includes starting with a clean polysilicon substrate and depositing a clean lead film on the oxide gate insulator at about 25°C to 75°C in a clean vacuum system. In one embodiment, the base pressure is approximately  $10^{-8}$  Torr or lower. The Pb film will have a thickness within 1 – 2 Å of its target value.

**[0038]** In one embodiment, lead and other metal films are deposited by a physical sputtering process. The sputtering process offers the ability to produce smoother films by increasing the re-sputtering-to-deposition ratio since re-sputtering preferentially reduces geometric high points of the film.

**[0039]** A low temperature oxidation process is then used to grow an oxide film of self-limited thickness. In one embodiment, oxygen gas is introduced at the desired pressure in order to oxidize the lead in situ without an intervening exposure to ambient air. For a fixed oxygen pressure and temperature, the PbO thickness increases with log(time). Its thickness can be controlled via time or other parameters to within 0.10 Å as determined via in situ ellipsometric or ex situ measurements of Josephson tunneling currents. This control over tunnel current is due to the excellent control over PbO thickness that can be achieved by low temperature oxidation.

**[0040]** For example, increasing the oxidation time from 100 to 1,000 minutes at an oxygen pressure of 750 Torr at 25°C only raises the PbO thickness by 3 Å (e.g., from about 21 Å to 24 Å). Accordingly, controlling the oxidation time to within 1 out of a nominal 100 minute total oxidation time provides a thickness that is within 0.1 Å of 21 Å. The PbO has a highly stoichiometric composition throughout its thickness as evidenced from ellipsometry and the fact that the tunnel barrier heights are identical for Pb/PbO structures.

**[0041]** Next, the system is re-evacuated and the top lead electrode is deposited. This produces a tunnel structure having virtually identical tunnel barriers at both Pb/O interfaces. The temperature used to subsequently deposit the polysilicon control gate is not critical. The PbO is stable to over 500°C and thus introduces no temperature constraints on subsequent processes.



[0042] In another embodiment, Al/Al<sub>2</sub>O<sub>3</sub> structures can be formed where the oxide is grown by low temperature oxidation in molecular or plasma oxygen. Capacitance and tunnel measurements indicate that the Al<sub>2</sub>O<sub>3</sub> thickness increases with the log(oxidation time). This is similar to that found for Pb/PbO as well as other oxide/metal systems.

[0043] Additionally, tunnel currents for an Al<sub>2</sub>O<sub>3</sub> tunnel barrier are asymmetrical with somewhat larger currents flowing when electrons are injected from the Al/Al<sub>2</sub>O<sub>3</sub> interface that is developed during oxide growth. This asymmetry is due to a minor change in the composition of the growing oxide. There is a small concentration of excess metal in the Al<sub>2</sub>O<sub>3</sub>, the concentration of which diminishes as the oxide is grown thicker. The excess Al ions produce a space charge that lowers the tunnel barrier at the inner interface. The oxide composition at the outer Al/Al<sub>2</sub>O<sub>3</sub> contact is much more stoichiometric and thus has a higher tunnel barrier. In spite of this minor complication, Al/Al<sub>2</sub>O<sub>3</sub> tunneling barriers can be formed that produce predictable and highly controllable tunnel currents that can be ejected from either electrode. The magnitude of the currents is still primarily dominated by Al<sub>2</sub>O<sub>3</sub> thickness that can be controlled via the oxidation parameters.

[0044] In one embodiment of the present invention, Al<sub>2</sub>O<sub>3</sub> metal oxide dielectrics can be formed by first thermally oxidizing the aluminum. In other embodiments, the aluminum is plasma oxidized or other oxidation methods can be used. Since the melting point of aluminum is much higher than lead, the formation of the Al/Al<sub>2</sub>O<sub>3</sub> structures are typically simpler than that used for the above-described Pb/PbO junctions.

[0045] In the Al<sub>2</sub>O<sub>3</sub> metal dielectric process of the present invention, the aluminum is sputter deposited on an oxide or other insulator at a temperature in the range of approximately 25°C to 150°C. Due to thermodynamic forces, the micro-crystals of the face centered cubic (f.c.c.) aluminum will have a strong and desirable preferred orientation.

[0046] The aluminum is then oxidized in situ in molecular oxygen using temperature, pressure, and time to obtain the desired Al<sub>2</sub>O<sub>3</sub> thickness. As with the lead oxide, the thickness of the aluminum increases with log(time) and can be controlled via time at a fixed oxygen pressure and temperature to within 0.10 Å when averaged over a large number of aluminum grains that are present under the counter-electrode. The thickness of

the  $\text{Al}_2\text{O}_3$  can be easily changed from about 15 Å to 35 Å by using appropriate oxidation parametrics. The oxide will be amorphous and remain so until temperatures in excess of 400°C are reached. The initiation of re-crystallization and grain growth can be suppressed, if desired, by the addition of small amounts of glass forming elements (e.g., Si) without altering the growth kinetics or barrier heights significantly.

[0047] The system is then re-evacuated and a layer of aluminum is deposited over the oxidized  $\text{Al}_2\text{O}_3$  layer. Finally, the polysilicon control gate layer is formed, using conventional processes that are well known in the art, on the layer of aluminum.

[0048] In additional embodiments, single layers of  $\text{Ta}_2\text{O}_5$ ,  $\text{TiO}_2$ ,  $\text{ZrO}_2$ ,  $\text{Nb}_2\text{O}_5$  and similar transition metal oxides can be formed by low temperature oxidation of transition metal films in molecular and plasma oxygen. They can also be formed by RF sputtering in an oxygen plasma or using other methods.

[0049] These metals oxidize via logarithmic kinetics to reach thicknesses of a few angstroms to tens of thousands of angstroms in a temperature range of 100°C to 300°C. Excellent oxide barriers for Josephson tunnel devices can be formed by RF sputter etching these metals in an oxygen plasma.

[0050] Lower temperature oxidation approaches of the present invention differ considerably from Metal-Organic Chemical Vapor Deposition (MOCVD) processes that are used to produce transition metal oxides. The MOCVD films typically require high temperature oxidation treatments to remove carbon impurities, improve oxide stoichiometry, and produce re-crystallization. Such high temperature treatments might also cause unwanted interactions between the oxide and the underlying silicon and, thus, necessitate the introduction of interfacial barrier layers.

[0051] In the processes of the present invention, control over the properties of the various transition metal oxides is improved from the prior art due to the limited thicknesses (approximately 10 Å to 100 Å) of metal that precludes the formation of significant quantities of unwanted sub-oxide films. This is due to thermodynamic forces driving the oxide compositions to their most stable oxidized state. In one embodiment, the duplex oxide layers are still crystallized. Such treatments can be done by RTP and will be shorter

than those used on MOCVD and sputter deposited oxides since the stoichiometry and purity of the low temperature oxides need not be adjusted at high temperatures.

[0052] The above-described processes for low temperature oxidation of various metals are for purposes of illustration only. The present invention is not limited to any one process for low temperature oxidation.

[0053] The embodiments of the present invention might also employ low temperature oxidation and short thermal treatments in an inert ambient atmosphere at 700°C in order to form a range of perovskite oxide films from parent alloy films. The dielectric constants of crystallized perovskite oxides can be very large (i.e., 100 to 1000). The transition metal layers can be either pure metals or alloys and the transition metals have similar metallurgy to their oxides. In contrast, the parent alloy films that can be converted to a perovskite oxide are typically comprised of metals having widely different chemical reactivities with oxygen and other common gasses.

[0054] If an alloy is to be completely oxidized, then thin film barriers such as Pd, Pt, or their conductive oxides should be added between the silicon and the parent metal film to serve as electrical contact layers, diffusion barriers, and oxidation stops. If the perovskite parent alloy film is only partially oxidized and covered with a second layer of the parent alloy, then the barrier heights will represent that developed during oxide growth at the parent perovskite alloy/perovskite oxide interface.

[0055] Figure 3 illustrates a functional block diagram of a memory device 300 that can incorporate the flash memory cells of the present invention. The memory device 300 is coupled to a processor 310. The processor 310 may be a microprocessor or some other type of controlling circuitry. The memory device 300 and the processor 310 form part of an electronic system 320. The memory device 300 has been simplified to focus on features of the memory that are helpful in understanding the present invention.

[0056] The memory device includes an array of the flash memory cells 330 that can be flash memory cells incorporating metal floating gates and high-k inter-gate dielectrics formed by low temperature oxidation of metals. The memory array 330 is arranged in banks of rows and columns. The control gates of each row of memory cells is coupled with

a wordline while the drain and source connections of the memory cells are coupled to bitlines. As is well known in the art, the connection of the cells to the bitlines depends on whether the array is a NAND architecture or a NOR architecture.

[0057] An address buffer circuit 340 is provided to latch address signals provided on address input connections A0-Ax 342. Address signals are received and decoded by a row decoder 344 and a column decoder 346 to access the memory array 330. It will be appreciated by those skilled in the art, with the benefit of the present description, that the number of address input connections depends on the density and architecture of the memory array 330. That is, the number of addresses increases with both increased memory cell counts and increased bank and block counts.

[0058] The memory device 300 reads data in the memory array 330 by sensing voltage or current changes in the memory array columns using sense/buffer circuitry 350. The sense/buffer circuitry, in one embodiment, is coupled to read and latch a row of data from the memory array 330. Data input and output buffer circuitry 360 is included for bi-directional data communication over a plurality of data connections 362 with the controller 310. Write circuitry 355 is provided to write data to the memory array.

[0059] Control circuitry 370 decodes signals provided on control connections 372 from the processor 310. These signals are used to control the operations on the memory array 330, including data read, data write, and erase operations. The control circuitry 370 may be a state machine, a sequencer, or some other type of controller.

[0060] The flash memory device illustrated in Figure 3 has been simplified to facilitate a basic understanding of the features of the memory. A more detailed understanding of internal circuitry and functions of flash memories are known to those skilled in the art.

## **CONCLUSION**

[0061] In summary, a flash memory cell transistor is comprised of a metal floating gate and a high-k inter-gate dielectric formed by low temperature oxidation of metals. Such construction provides thinner floating gates and, therefore, higher coupling capacitances between the floating gate and control gate and between the floating gate and substrate. The

low-k dielectric isolation material between the array rows provides low stray capacitances for row-to-row floating gate capacitances and end-to-end capacitances.

**[0062]** Although specific embodiments have been illustrated and described herein, it will be appreciated by those of ordinary skill in the art that any arrangement that is calculated to achieve the same purpose may be substituted for the specific embodiments shown. Many adaptations of the invention will be apparent to those of ordinary skill in the art.

Accordingly, this application is intended to cover any adaptations or variations of the invention. It is manifestly intended that this invention be limited only by the following claims and equivalents thereof.